Eighth USA/Europe Air Traffic Management Research and Development Seminar (ATM2009)

Identification of Robust Routes using Convective Weather Forecasts

Diana Michalek and Hamsa Balakrishnan
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA
{dianam, hamsa}@mit.edu

***Abstract*– Convective weather is responsible for large delays and widespread disruptions in the U.S. National Airspace System (NAS), especially during summer months when travel demand is high. This has been the motivation for Air Traffic Flow Management (ATFM) algorithms that optimize flight routes in the presence of reduced airspace and airport capacities. These models assume either the availability of reliable probabilistic weather forecasts or accurate predictions of robust routes; unfortunately, such forecasts do not currently exist. This paper adopts a data-driven approach that identifies robust routes and derives stochastic capacity forecasts from deterministic convective weather forecasts. Using techniques from machine learning and extensive data sets of forecast and observed convective weather, the proposed approach classifies routes that are likely to be viable in reality. The resultant model for route robustness can also be mapped into probabilistic airspace capacity forecasts.**

Keywords- convective weather; air traffic management; integration of weather forecasts and air traffic management; route robustness; airspace capacity

I. INTRODUCTION

The increase in demand for air travel over the past few years has been accompanied by an increase in congestion and delays in the National Airspace System (NAS) of the United States, and has made the system more susceptible to weather disruptions. This problem is particularly intense during summer months, when travel demand is high and there are frequent thunderstorms (convective weather activity) over much of the continental U.S. It has been estimated by the Joint Economic Committee of the U.S. Senate that domestic air traffic delays in 2007 cost the U.S. economy $41 billion [1]. It has also been estimated that 76.9% of all delay in the NAS and 25% of all delayed flights in 2007 was weather-related [1, 2]. With the demand for air traffic operations expected to grow significantly over the next two decades, it has become increasingly important to develop approaches that will enable the efficient operation of the airspace system, even in the presence of convective weather [3].

A. Background and related work

There has been much research over the past several decades on techniques to minimize air traffic delays and to better balance demand for air traffic operations and the available capacities of various airspace and airport resources. This research falls broadly into the realm of Air Traffic Flow Management (ATFM), which is the process of making strategic decisions a few hours ahead of the time of operations, in order to balance the demand for, and capacity of, constrained NAS resources. However, the capacity of airspace resources is strongly influenced by ambient weather, since aircraft need to avoid hazardous atmospheric conditions and may therefore be forced to deviate from their planned trajectories. Traditionally, ATFM models handled the presence of weather by assuming that the impact of weather on the capacity of a resource at any time was known, and used the deterministic estimates of capacity to route flights between their origins and destinations in order to minimize delays. Various approaches have been adopted to solve the large scale optimization problems that arise, including integer programming formulations [4] and Eulerian models which treat the traffic as continuous flows [5, 6]. Algorithms have also been developed to efficiently synthesize routes through regions of airspace impacted by convective weather. These algorithms require fine-grained and time-varying weather forecast data as static weather input, and focus on synthesizing short and easily flyable routes which do not get too close to regions of airspace impacted by weather [7, 8]. The challenge in using these deterministic approaches lies in the fact that under clear weather conditions, deterministic capacity estimates based on weather forecasts tend to be stable and tend to reflect the conditions that materialize; however, under stormy weather conditions, capacity is highly variable and the use of the expected capacity for planning is unrealistic.

The knowledge that weather forecasts are inherently uncertain motivated optimization approaches that assumed multiple capacity scenarios for airspace resources, with associated probabilities of occurrence. These approaches then minimized the expected value of delay in the system while trying to route aircraft so as to not violate capacity constraints [9]. More recently, robust optimization approaches have been proposed that assume a set of possible capacity uncertainty values, and try to keep the system safe for any possible realizations of the uncertainty [10]. At the tactical level, prior research has assumed that convective weather can be modeled as a dynamic stochastic process, and flight routes determined using dynamic programming.

There have also been recent attempts at the problem of creat-

ing stochastic and deterministic models of capacity from weather forecasts. In [11], the authors considered the problem of estimating the capacity of a sector of en-route airspace by computing a theoretical capacity given weather in the region. This was done through the application of continuous maximum flow theory. This work relied on static weather forecasts and did not incorporate uncertainty intervals or any measure of forecast accuracy. In [12], the authors extended this approach to the case of weather forecasts accompanied by regions of uncertainty. However, the uncertainty profiles were randomly generated, by assuming that the probability of a weather impacted region of a particular size was proportional to the intensity of the weather forecast in the region. In the terminal area environment, the Route Availability Planning Tool (RAPT) uses Lincoln Lab Convective Weather Forecasts to model jet route blockage deterministically. The product is used operationally in the New York area airports to help controllers determine if aircraft can take off over relatively short time horizons [13].

B. *Contributions of this paper*

While the efforts described above assume the existence of reliable probabilistic weather (or capacity) forecasts, no attempts have been made to evaluate the quality of existing forecast products nor their predictions. Instead, the forecasts have been treated as ground truth. In contrast, this paper explicitly considers the problem of understanding and validating weather forecasts, and developing techniques that will help integrate them into ATM decision-making in a reliable and meaningful fashion. We adopt a data-driven approach to achieving this objective. We make use of state-of-the-art aviation convective weather forecasts, developed by MIT Lincoln Laboratory, to identify robust routes, that is, routes that are likely to remain viable in the actual weather that materializes. We consider various features (characteristics) of the forecast weather along arrival and departure routes, and identify features which are highly correlated with route blockage. Using techniques from machine learning, we propose potential classification algorithms that predict whether a given route is likely to be open or blocked in actual weather, based on the values of different features of the route, as determined by the forecast. We compare these techniques with each other as well as the naive prediction (which would treat the forecast as ground truth, and classify a route as blocked if it is blocked in the forecast weather). We evaluate these different approaches using several metrics, such as the accuracy (the fraction of time that the prediction is correct), the false positive rate (the fraction of time that we forecast that the route will be open but it ends up being closed), the false negative rate (the fraction of the time that we forecast that the route will be closed, but instead it remains viable), etc.

In prior work, we considered the problem of evaluating forecasts using pixel-by-pixel comparisons, and also evaluated the probability that a route that is open in the forecast remains open, by considering features individually [14]. In contrast, this paper adopts a route-centric view: we analyze the probability that any given route remains open in the observed weather, by developing classification algorithms that consider combinations of features.

II. Problem Description

In this section we formalize the problem of identifying robust routes in the terminal area. We also introduce the Lincoln Lab Convective Weather Forecast (CWF) and the dynamic forecast grid, used in constructing our route robustness model. For simplicity, we consider the case of airports with well-defined arrival and departure gates through which most aircraft are routed. An instance of such an airport is Hartsfield Atlanta International airport (ATL), which uses four arrival gates in the NE, NW, SE and SW corners, and four departure gates in the North, South, East and West corners.

A. *Terminal-area model*

Consider the following version of the route robustness problem, illustrated in Figure 1. The input is a terminal-area, defined by two concentric circles: an outer circle C_O of radius R, and an inner circle C_I of radius r. The outer circle C_O represents the points at which arriving aircraft first enter the terminal airspace, and R is typically about 40 nm (75 km). The inner circle C_I represents the point at which aircraft start their final approach into the airport, and is assumed to be 10 km in this study. In contrast, departures traverse the terminal-area in the reverse direction, entering it close to the airport at C_I and exiting it through the outer boundary C_O.

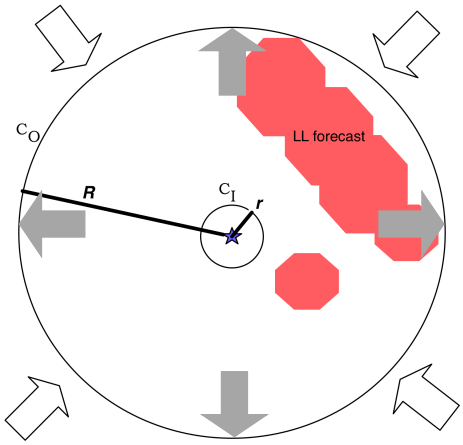

Figure 1: Model of terminal-area flows. Arrival flows enters through the outer circle C_O and flow into the inner circle C_I, while departure flows (denoted by grey arrows) travel in the reverse direction. The red region represents a forecast weather hazard.

Given a route (for example, a path between an arrival gate on C_O and a point on C_I, a weather forecast provides us with a prediction of where the weather obstacles will be located, and therefore a prediction of whether the route will remain or not. However, we note that weather forecasts are not always accurate. Figure 2 (left) shows an illustrative example: three paths overlaid on a 30-min weather forecast on the left; and the same paths overlaid on the observed weather for that scenario on the right. We notice that two of the three paths (denoted by blue lines) are predicted to be open but are blocked by weather in reality, while the third (denoted by a red line) is forecast to be blocked, but is open in the weather that actually materializes.

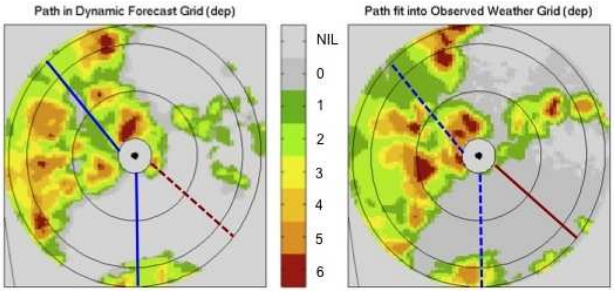

Figure 2: An example of forecast inaccuracy. The figure on the left shows the forecast, while the corresponding observed weather is on the right. We note that two of the paths are forecast to be open but are blocked in reality, while a third is forecast to be blocked but is open in reality.

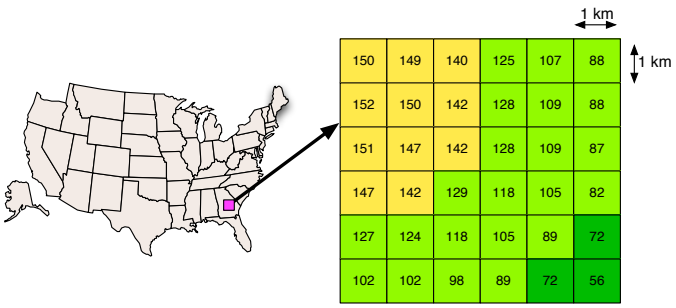

Figure 3: Sample Lincoln Lab Convective Weather Forecast near ATL.

B. *Problem statement*

The objective of this paper is to determine routes that are likely to be robust to weather disruptions, by understanding and incorporating the inherent uncertainty associated with weather forecasts. In other words, our problem can be stated as follows:
Given a weather forecast for some time in the future and a set of predetermined potential routes, we would like to best identify those routes that are likely to be open in the actual weather that materializes and also quantify the uncertainty associated with our prediction mechanism.

In the approach that we follow to solve this problem, we use the following definition for an open route.

Definition 1 *A route is defined to be **open** or **clear** in the observed weather if there exists a route that is not impacted by weather within a small neighborhood of the original route.*

This relaxed definition allows for slight deviations in a planned route that reflect the "wiggle room" or the ability of an aircraft to make small adjustments to the planned route.

The problem stated above is an important one from the air traffic management perspective for several reasons. First, it aims to capture trends in how the impact of observed weather on routes differs from predicted impact, rather than by simply evaluating forecasts using pixel-by-pixel comparisons [14]. Second, it takes into account the realities of scheduling aircraft routes, such as the ability to allow small deviations from planned routes without effecting operations. Third, this approach suggests that in the terminal-area, the theoretical capacity may not be a sufficient metric to measure the impact of weather on air traffic flows [11]. This is because while the theoretical capacity might predict that N aircraft will be able to enter airspace over the next hour, it may not indicate the possibility (which is critical for planning) that these aircraft must necessarily arrive from the West. Furthermore, it is possible for the forecast theoretical capacity to exactly match the realized theoretical capacity, and yet require that aircraft use trajectories that are very far from the original planned routes.

C. *Lincoln Lab's Convective Weather Forecast*

In order to assess the robustness of a route to the differences between the forecast and actual weather, it is necessary to first select a weather forecast. This paper uses the state-of-the-art Lincoln Lab Convective Weather Forecast (CWF), which is briefly described in this section.

The 0-2 hour CWF consists of a grid of 1km $\times$ 1km pixels covering a large portion of the NAS [15]. Each pixel contains a predicted value of Vertically Integrated Liquid (VIL), indicated by an integer value in the range $[0,255]$. Figure 3 shows a sample forecast for ATL. These VIL values are divided into seven levels of convective activity, ranging from level 0 (none) to level 6 (very severe). A VIL value above a certain threshold (133, in practice) in the observed data corresponds to weather of severity level 3 or higher, which is commonly considered to be hazardous to pilots. A forecast has a horizon that spans every 5 minute increment between 5 and 120 minutes, and is updated every 5 minutes. In other words, at time T_0, forecasts are available for time $T_0+5, T_0+10, T_0+15, \ldots, T_0+120$. The forecast data is accompanied by observed VIL values for the same region of airspace at that time, providing data that can be used for evaluating the quality of the forecast.

The static CWF is useful in obtaining a general idea of what weather will look like, and is used in various decision support tools by air traffic controllers and airlines. Lincoln Lab, as well as other entities that develop forecast products, provide daily statistics such as rates of false positives, false negatives, and a skill score, but these are pixel-based, and often ad hoc. It is important to note that no large-scale historical evaluation of forecast accuracy for ATM decision-making has been performed so far.

D. *Dynamic weather grid*

To model aircraft moving through the terminal area, it is necessary to use a different time horizon for different aircraft positions. We achieve this dynamic weather grid by splicing together weather data for time instants t that increase from the outer to inner circle for arrivals, and decrease from the outer to inner circle for departures. The distance between two concentric circles in the grid (shown in Figure 4) corresponds to the distance flown by a typical aircraft in 5 min. These circles are drawn assuming an average aircraft speed of 180 knots in the terminal area; we have also conducted similar analysis for aircraft speeds of 85 knots (corresponding to slower, general aviation aircraft).

Figure 4 contains a sample dynamic weather grid for arriving aircraft. We assume aircraft arrive at C_O at time t, with a

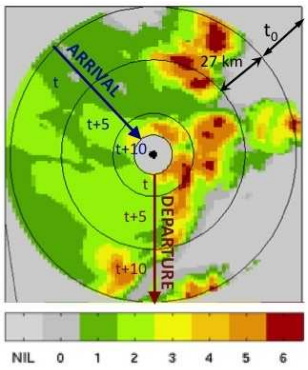

Figure 4: Sample forecast region for arrivals, created by splicing together consecutive 5-minute forecasts. This is for a 60-minute time horizon on June 8, 2007, where aircraft reach the outer circle at time 2130 hours.

t_0-minute time horizon. For departures, the corresponding dynamic grid assumes aircraft arrive at C_I at time t, with the same t_0-minute time horizon. This grid will therefore be used for planning at the current time, namely, time $t - t_0$.

III. Generation of Data Sets

As has been mentioned before, this paper adopts a data-driven approach of identifying routes that are likely to be robust to the inaccuracies in the forecast. The approach is based on a large-scale evaluation of the performance of the Convective Weather Forecast, and the difference in predicted and observed impact on routes. An essential step is therefore the generation of the necessary data sets, which consists of the selection of forecast and observed weather scenarios, selection of potential arrival and departure routes, and the validation of these routes in observed weather, as described in this section.

A. Selection of weather scenarios

A dataset was created containing routes for several weather scenarios during each of the 14 most weather-impacted days in ATL during the months of June and July 2007, when ranked according to weather-related delays.

Although the Lincoln Lab CWF data can be described as a matrix of integers in the range $[0,255]$, the archives of this data are kept in a proprietary format, and each day of data takes several hours to extract, yielding 30 GB of uncompressed binary data. To identify convective weather scenarios for the ATL terminal-area, the forecasts for the airspace surrounding ATL were extracted and visualized to identify the time periods with maximum convective weather activity. This resulted in an average of 4 weather scenarios per day, separated from each other by at least 30 minutes, and yielded a total of approximately 400 trajectories in forecast weather. Ten datasets were created, corresponding to the 10-, 30-, 60-, 90- and 100-minute time horizons for both departures and arrivals.

B. Route selection in the forecast grid

Potential aircraft trajectories through the forecast grid of each weather scenario are generated by sampling eight straight routes from C_0 to C_I, as depicted in Figure 5. These eight trajectories represent a sampling of routes through varying weather forecasts. Arrival trajectories point toward the inner circle, while the departure trajectories are oriented in the opposite direction.

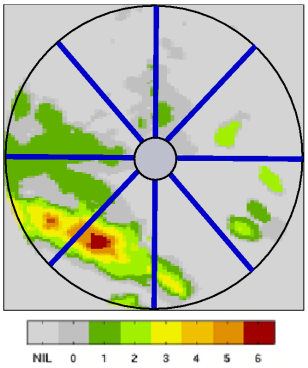

Figure 5: Eight routes selected through a 60-minute departure forecast scenario over ATL for June 12 2007, at 0600 hours.

C. Validation of routes in the observed weather grid

Each path P generated in the manner described above is evaluated using the observed weather data. A route P is defined as *open* if there exists a corresponding route in the observed weather grid which is within B km of P and does not pass through any actual weather hazards. This B-km neighborhood allows for slight perturbations in the path (on the order of several kilometers), which represents only a slight change from the original planned trajectory, P.

Open routes are synthesized by solving the following modified shortest-path problem through the dynamic grid of observed weather:

Construct a directed graph $G(\mathcal{N},\mathcal{A})$ such that the set of nodes $\mathcal{N}$ contains all pixels within B km of P (in the dynamic observed weather grid) which are free of weather hazards, and such that each set of adjacent nodes form an arc $a \in \mathcal{A}$. At time t, a unit of flow is sent from a set of source nodes $\mathcal{S} = C_O \cap \mathcal{N}$ (the subset of nodes lying on the outer circle C_O) to a set of sink nodes $\mathcal{T} = C_I \cap \mathcal{N}$. For simplicity, we use a standard transformation and introduce a supersource $\mathcal{S}$ and a supersink $\mathcal{T}$, and route one unit of flow between the two through the source nodes and sink nodes [16]. Define $\mathrm{NX}(i,j)$ to be the node $k \in \mathcal{N}$ which constitutes a straight next arc if (i,j) is used. In other words, nodes i,j,k form a straight line in the observed weather grid. The objective is to find the minimum cost flow f such that out of all minimum cost flows, f has the minimum number of turns.

This problem is modeled by the IP below, which is a slight modification to the shortest path problem. This problem is solved for each of the selected routes in the data set; the infeasibility of the problem implies that the route is blocked in the observed weather grid, feasibility implies that the route is considered open. A version of this problem can be also be solved with different sets of sources and sinks to generate a large set of candidate paths for a given weather forecast scenario [14]. Furthermore, although the construction above models the case of arrivals, the exact same IP can be used to model departures as well, as long

as the underlying dynamic grid is changed.

$$x_{ij} := \text{flow on arc } (i,j) \in \mathscr{A}$$
$$z_{ij} := 1 \text{ if } (i,j) \in \mathscr{A} \text{ is a turn}, 0 \text{ otherwise.}$$

$$\min \sum_{(i,j)\in\mathscr{A}} c_{ij}x_{ij} + \lambda \sum_{(i,j)\in\mathscr{A}} z_{ij}$$

$$\text{s.t.} \sum_{\substack{j\in\mathscr{N}:\\(i,j)\in\mathscr{A}}} x_{ij} - \sum_{\substack{j\in\mathscr{N}:\\(j,i)\in\mathscr{A}}} x_{ji} = b_i \quad \forall i \in \mathscr{N} \tag{1}$$

$$z_{ij} \geq x_{ij} - \sum_{\substack{k\in NX(i,j):\\(j,k)\in A}} x_{jk} \quad \forall (i,j) \in \mathscr{A} \tag{2}$$

$$x \in \{0,1\}^n \tag{3}$$

$$z \in \{0,1\}^n \tag{4}$$

Constraints (1) are the flow balance constraints, with $b_i := -1$ for a supersource $\bar{\mathscr{S}}$, $b_i := +1$ for a supersink $\bar{\mathscr{T}}$, and $b_i := 0$ for all other nodes i in $\mathscr{N}$. Constraints (2) in conjunction with the penalty term in the objective function serve to minimize the number of turns in the path without changing the path length, since it is desirable that aircraft trajectories have a limited number of turns for simplicity. All arcs that follow (i,j), *except* (j,k) for $k = \text{NX}(i,j)$, pay a penalty in the objective function. λ is chosen to be sufficiently small (less than the maximum length of any path) to ensure that a longer route with fewer turns is never chosen. Finally, x and z are binary variables because a single path cannot be split up, and the existence of a turn is a binary quality.

D. Dataset Details

The overall statistics of the route blockage datasets for arrivals and departures at the five time horizons studied are listed in Table 1. Each dataset contains approximately 400 routes, the majority of which are open. The percentages of open forecast routes (routes which do not pass through Level 3+ weather in the dynamic forecast grid) are between 50 and 66 percent for both arrivals and departures, meaning that approximately half of the routes in the dataset are forecast to be blocked. However, these same routes are open over 78% of the time in the weather that materializes (that is, there is a route in the neighborhood of the original path which does not pass through Level 3+ weather in the dynamic observed weather grid). The last two columns indicate how the forecasts and true weather differ for individual routes. Routes that are forecast as open are overwhelmingly open in the observed weather grid, with rates of 87% and above. Arrivals have slightly lower rates than departures, and the rates decrease with increasing time horizon. Both of these trends are to be expected, because arrivals typically encounter the bottleneck at the end of their route through terminal airspace, where the forecasts are less accurate. Finally, routes that are forecast as closed are closed in the true weather approximately 60% of the time. These low rates reflect the effect of the additional flexibility allowed for finding routes in the actual weather. Figure 6 contains examples of routes synthesized in the forecast grid, along with the same routes validated against the observed weather.

	t_0	# Paths	Fx Open (%)	Act. Open (%)	% Act. Open | Fx Open	% Act. Closed | Fx Closed
Arrivals	10	408	51	78	99	57
	30	408	54	78	96	58
	60	384	55	78	93	60
	90	392	63	78	88	62
	100	408	65	78	87	62
Departures	10	408	55	79	99	55
	30	408	53	79	95	61
	60	384	55	79	94	62
	90	392	61	79	91	60
	100	408	66	79	90	58

Table 1: Overall dataset statistics for each of the 8 datasets. Fx Open (Actual Open) refers to the percent of routes that are open in the forecast (actual) weather grid. [Act. open | Fx Open] refers to the percentage of forecast open routes which are open in the actual weather as well. [Act. closed | Fx closed] has the similar connotation for closed routes.

The raw data suggest that subject to minor adjustments, planning at a 10-, 30-, 60-, 90-, and 100-minute time horizons is quite reasonable, since routes that are forecast to be open end up being overwhelmingly so. This is encouraging, and shows that allowing even small adjustments from fixed arrival routes can improve the quality of decision-making based on the forecast. The next sections explore how we can learn more from these data sets, and better predict blockage based on the forecast data.

IV. Feature selection

Once a dataset of routes through the weather-constrained terminal-area is available, it is interesting to identify characteristics of the convective forecast which may best reflect the likelihood that a trajectory will be opn in the observed weather.

A. Potential features of interest

For each path, eleven features of interest were indentified and each feature was correlated with route blockage. The eleven features of the forecast weather, chosen for their possible correlation with route blockage, are listed below:

1. Mean VIL along the path
2. Standard Deviation of VIL along the path
3. Minimum distance to level 3+ weather along the path
4. Mean distance to level 3+ weather along the path
5. Maximum VIL in neighborhood of the path
6. Theoretical capacity for the weather scenario
7. Number of segments in the minimum cut
8. Length of the minimum cut segment (bottleneck) that the path passes through
9. Length of tightest bottleneck
10. Maximum pixel density of L3+ weather along path
11. Maximum VIL density along path

The first four features are reasonably self-explanatory, but the others require some explanation. Feature 5 is the maximum VIL forecast in the neighborhood of radius B along the path, where B is the same as in the integer program in Section II. Features 6,7, and 8 refer to the theoretical capacity of the forecast grid and

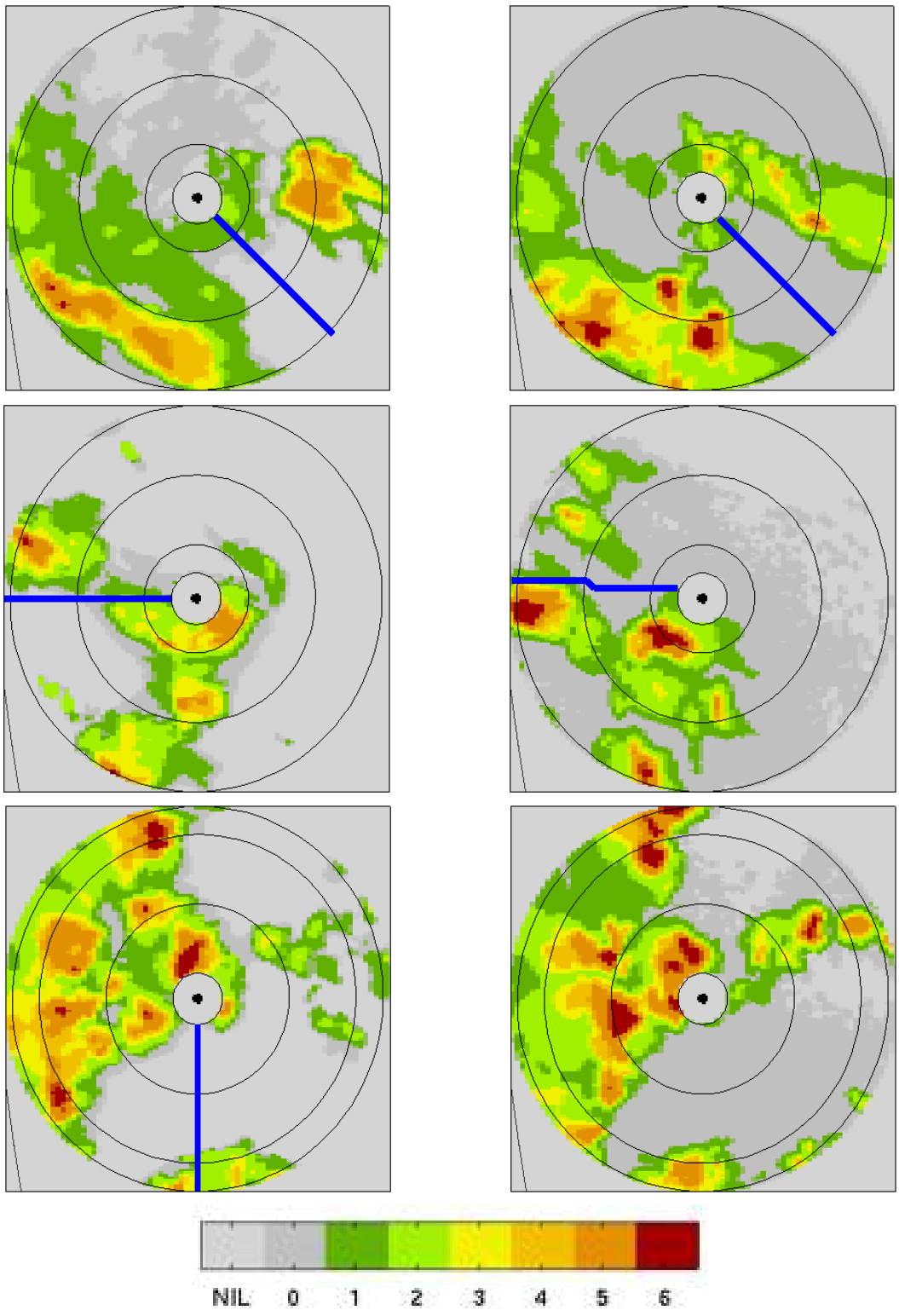

Figure 6: Sample routes in a dynamic forecast grid are on the left-hand column, and the corresponding routes in the actual weather are on the right-hand column. The top weather scenario is an arrival route from June 12, 2007 at 0630hrs with a 60-minute time horizon, and depicts a situation where the route that is open according to the forecast ends up open in the actual weather that materialized. The middle scenario shows an arrival route from June 15, 2007 at 2000hrs with a 90-minute time horizon. The precise forecast route is blocked according to the forecast, but a nearby route is available in the true weather grid. The bottom scenario is a departure route from June 8, 2007 at 2030hrs with a 30-minute time horizon. In this situation, the forecast route is not open in the observed weather grid.

the corresponding minimum cut, and are computed using continuous max flow theory and the techniques described in [17, 11]. Feature 9 contains the length of the minimum bottleneck through which the route passes. Feature 10 is meant to indicate the intensity of the weather in the neighborhood of the route. It is computed by taking a B km neighborhood of the route, and finding the strip of pixels perpendicular to the route with the largest percentage of Level 3+ forecast pixels. If the route is forecast to pass through Level 3+ weather, Features 8-9 will be 0, but Feature 10 may still contain pertinent information about the nature of the weather through which the route passes. Finally, Feature 11 is computed similar to Feature 10, except that it considers the largest average VIL in a perpendicular strip (rather than the largest percentage of Level 3+ weather).

B. Feature selection

Previous work by the authors computed the simple correlations for each feature with blockage, giving smooth estimates of the probability of blockage at each feature level [14]. To evaluate features for classification and gain a better understanding of which features best correlate with blockage individually, we compute the Mutual Information between each feature X_i and the blockage label y (+1 for open, -1 for blocked).

Mutual information is an information-theoretic measure of the dependence between two random variables X and Y, and measures how much the uncertainty of X is reduced if Y is observed. Note that this measure considers each feature individually and does not capture situations in which two random variables combined correlate very well with y. For discrete random variables X and Y, their mutual information, $I[X;Y]$, can be expressed as

$$I[X;Y] = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}$$

To compute mutual information, it is necessary to have access to the density functions for the corresponding random variables. When the dataset size is much larger than the range size of the joint p.d.f $F_{X,Y}$, we can choose the Maximum Likelihood parameter estimates of the p.d.f.s as good approximations. For the case of continuous random variables, the data are discretized by placing points into k equally sized bins. We note that there are other approaches to approximating MI for continuous distributions, involving setting bin sizes so that the data points are equally distributed between the bins, which is a better approximation to true entropy [18, 19]; however, for simplicity, these methods are not adopted here.

Figure 7 contains a comparison of mutual information (MI) across features and time horizons for both departure and arrival datasets. It is seen that MI decreases overall as the time horizon increases, which reflects the decreased forecast accuracy at longer time horizons. In addition, departures have slightly higher MI than arrivals across the board, which can be explained by the fact (also discussed in Section III.D) that departures enter the bottleneck of their path (close to the inner circle) at the start of their time through the terminal area. Features 1, 10 and 11 consistently have the highest MI scores, while features 6 and 7 have the lowest.

The above analysis provides a better understanding of how well the features of a convective weather forecast correlate with route blockage. In the next section, the selected features will be used to predict robust routes, though the use of methods from machine learning.

V. Classification

In this section, using the route datasets described in sections III and IV, techniques from machine learning are adapted to better predict the possibility of route blockage in actual weather. Specifically, a classifier is trained to predict, given the features of a route in forecast weather, whether the route will be open or blocked in the actual weather that materializes. This prediction is also associated with a probability, which is determined by the performance metrics of the classifier.

A. Training objectives

When evaluating a classifier, the class predications are compared with the actual classes of a test set, according to the standard *two-class confusion matrix*:

	Predicted Open	Predicted Blocked
Actual Open	TP (True Positive)	FN (False Negative)
Actual Blocked	FP (False Positive)	TN (True Negative)

Although it is typically desirable to maximize the accuracy (total correctly predicted items) of a classifier on a test set, the setting of aviation weather warrants a modified objective. Due to safety concerns, it is more important to correctly predict a route that ends up blocked than one that ends up open. This emphasis on correctly predicting members of the blocked class (minimizing false positives) is complicated by the fact that the dataset is imbalanced, having fewer blocked examples than open, making it inherently harder to perform well on the minority class.

In addition to the FP and FN rate, we compute the following (standard) performance metrics to the evaluate our classifier: $a^- = \frac{TN}{TN+FP}$, $a^+ = \frac{TP}{TP+FN}$, g-mean = $\sqrt{a^- * a^+}$, and accuracy = $\frac{TP+TN}{n}$, where n is the total number of routes in the data set. a^- (also known as recall) is a measure of how well the classifier performs on members of the blocked (minority) class. We will seek to maximize this value through classification.

B. Two ensemble classifiers

The Machine Learning literature has shown that ensemble classifiers tend to perform well on imbalanced datasets, outperforming non-ensemble methods [20, 21]. We trained two classifiers using the R language for statistical computing along the lines of [22]: an Ensemble of Support Vector Machines (EnsSVM), and a weighted random forest (WRF). This section describes the training process.

For both classifiers, we created identical training and test data sets from each base route dataset. We partitioned the base dataset randomly so that the training set had 70% of instances, and the test set 30%, making sure that weather scenarios from the same date were not split up, so as not to introduce bias. The training

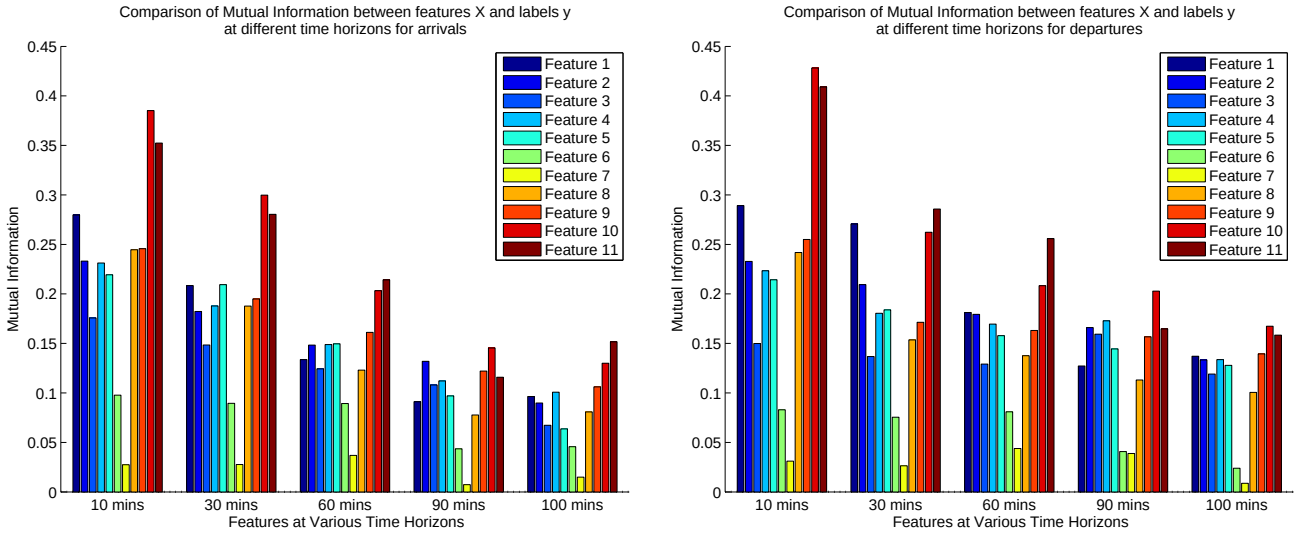

Figure 7: Comparison of mutual information values across features and time horizons for both arrivals (left) and departures (right)

set was then further processed when setting up the ensemble: m blocked instances of the training set were set aside, and N bootstrap samples of size m were created from the open instances. The blocked set was then combined with each of the bootstrap samples to create N training bootstrap training sets. This way, each of the N bootstrap sets had a balance between open and blocked instances.

These N bootstrap samples were then used to train the two types of classifiers. EnsSVM was trained with an RBF kernel, and 5-fold cross validation was used to tune the parameters. The WRF was trained using the rpart package for R [23] for a large set of weights, where a higher weight increases the penalty for missclassifying blocked examples.

In both cases, the resulting ensemble classifier uses the majority vote of the ensemble to classify new routes.

C. *Results for Ensemble SVM*

Table 2 shows the results for the ensemble SVM classifier at all four time horizons of interest, for both arrivals and departures. All metrics shown are the average of 5 runs of the classifier (on independently generated test/training sets), to account for variability in training.

The table illustrates two major trends. At the shorter time horizons of 10-, 30-, and 60-minutes, the ensemble does not improve the performance of the forecast on blocked routes, since the recall rates (a^-) of the classifier and forecast are quite similar, although there is improvement in overall accuracy. This is not surprising since at these short time horizons, there is very little for improvement, and the weather forecasts are known to be more accurate.

There is clear improvement in the recall rate of the classifier at 90- and 100- minute time horizons. Arrivals at 90-minutes post a 18% improvement in recall rate over the weather forecast, at a similar cost to FP rate. At 100-minutes, the improvement in recall rate is 37%, with a slightly larger cost to accuracy. The results for departures show the same trends.

Thus EnsSVM is successful in combining the features of a given weather scenario and using them to predict route blockage, with higher recall rates than the weather forecast at longer time horizons. This decrease in the false positive rate comes at an expected tradeoff with accuracy, due to the conservative objective function we placed on the learning algorithm, and the imbalance between open and blocked routes.

D. *Results for the weighted random forest*

The performance of the WRF classifier is similar to that of EnsSVM, as it is successful in learning from the features to predict blocked routes, at a cost to overall accuracy. Although the associated metrics are omitted due to space constraints, the explicit penalty on misclassifying blocked routes in the WRF (in the form of a weight in the training loss function), provides an illustration of the tradeoff between FP rate and accuracy.

Figure 8 depicts this relationship across four time horizons. A diagonal trend is evident between the FP rate and the accuracy rate of the WRF for each time horizon. The label on each point contains the weight used in the training function. Points associated with a lower weight tend to be in the top right (higher FP rate and accuracy), while points associated with a higher weight tend to be in the bottom left (lower FP rate and accuracy), for each time horizon. The figure also depicts the changes in accuracy and FP rates across the time horizons: at the shorter time horizons, the classifier can attain higher accuracy rates and lower FP rates (due to the greater reliability of the weather forecast), while at the longer time horizons, the absolute improvement in FP rate is greater, but at a correspondingly larger cost to accuracy, than at shorter time horizons.

Two additional classifiers, namely, a (regular) SVM with an RBF kernel and a decision tree with a weighted loss function, were trained on the route blockage data set, in order to validate the results above and compare with other classification methods.

		10-min		30-min		60-min		90-min		100-min	
		EnSVM	Fx	EnSVM	Fx	EnSVM	Fx	EnSVM	Fx	EnSVM	Fx
Arrivals	Acc	81.77	74.22	75.57	71.4	71.64	69.86	52.76	70.31	28.66	69.01
	a^-	89.95	96.97	86.17	90.6	71.28	72.15	87.9	69.55	93.71	56.22
	a^+	79.22	68.43	71.49	65.14	70.95	68.55	42.38	70.44	9.25	72.28
	g-mean	0.84	0.81	0.78	0.77	0.71	0.7	0.53	0.70	0.29	0.63
	% TP	64.15	55.32	55.99	50.95	57.43	55.48	32.66	54.30	7.16	55.84
	% FP	1.72	0.45	3.16	2.29	5.18	5.00	2.73	6.82	1.34	9.66
	% TN	17.62	18.89	19.59	20.46	14.20	14.38	20.1	16.01	21.49	13.17
	% FN	16.5	25.33	21.27	26.30	23.19	25.14	44.51	22.87	70.00	21.33
Departures	Acc	77.3	71.97	78.2	72.27	68.94	69.29	72.92	76.07	52.37	73.11
	a^-	98.31	99.31	88.55	90.52	82.35	80.92	86.43	76.7	80.84	68.55
	a^+	69.91	62.55	73.91	66.1	64.83	65.6	69.66	75.64	45.26	74.34
	g-mean	0.83	0.79	0.8	0.77	0.73	0.73	0.77	0.76	0.52	0.71
	% TP	53.05	47.45	57.77	51.51	51.92	52.45	57.71	62.6	35.5	58.74
	% FP	0.44	0.17	2.32	1.99	3.42	3.6	2.39	4.12	4.13	6.63
	% TN	24.25	24.52	20.43	20.76	17.01	16.84	15.21	13.48	16.87	14.37
	% FN	22.26	27.86	19.48	25.74	27.64	27.11	24.69	19.81	43.49	20.25

Table 2: Results for (the average of 5 runs of) the Ensemble SVM classifier, for arrivals and departures

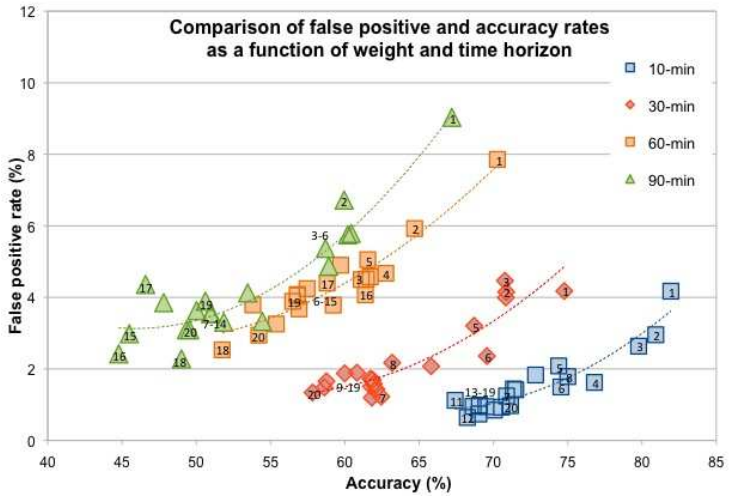

Figure 8: Comparison of false positive and accuracy rates of the weighted random forest classifier for each weight (where each point is the mean of 10 iterations). Each of the 4 trend lines (one for each time horizon) depicts the changes in classifier performance as a function of weight. In general, a lower false positive rate is accompanied by lower accuracy, and higher weight (penalty against false positives).

EnsSVM and WRF outperformed them in terms of maximizing recall. Due to space constraints we omit a discussion of these techniques.

VI. Capacity forecasts from predictions of route blockage

The route blockage model can be used to create a stochastic model of capacity. This section presents an initial version of such a model, for the case of arrivals. For an airport with m arrival gates (in the case of ATL, $m = 4$ as shown in Figure 9), and for a given time horizon t_0, we can forecast capacity in the following way. First, the four standard arrival routes are sampled, each sourced from a different quadrant of the outer circle C_O through the forecast grid. For each of these routes, the classification algorithm predicts the route to be either open or closed, and also provides an estimate of the probability with which the classifier believes that the route will be open. This probability can be used to represent the probability that the route will be blocked in the true weather grid given the EnsSVM prediction. Let C be the clear-weather capacity of the airspace. Then the capacity of the the airspace can be forecast as $\frac{Ck}{m}$ with probability Pr(exactly k of the arrival routes are open).

VII. Conclusion

This paper presents a data-driven approach to the prediction of routes that are likely to be robust to the inaccuracies of convective weather forecasts. In contrast to prior research in air traffic management which assumed the presence of accurate deterministic or probabilistic capacity forecasts as inputs, this approach evaluates features of weather forecasts, and selects ones that have high correlation with route blockage in observed weather. These features are then utilized in classification algorithms based on machine learning techniques to predict, given a set of potential routes and a weather forecast, which routes are likely to be blocked and which ones will be open in the observed weather. The performance of the proposed classifiers is evaluated and

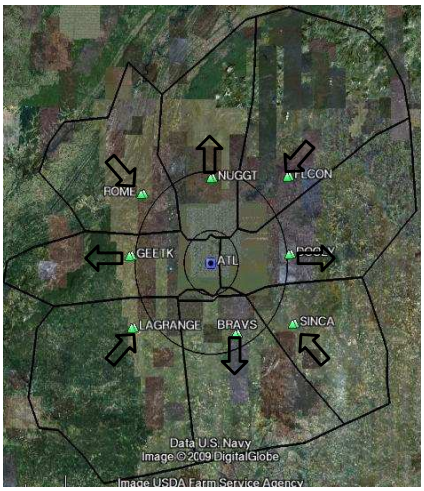

Figure 9: ATL terminal-area, showing the arrival and departure gates.

compared to the naive forecast predictions, using several metrics including the false positive rate (or FP rate, when a route is predicted to be open but is blocked in the weather that materializes), and the overall accuracy. It is shown that the classifiers can be optimized to minimize the FP rate, which is important for this application, and the tradeoffs between overall accuracy and the FP rate are illustrated. Finally, a possible approach to using these route robustness models to obtain probabilistic capacity forecasts is discussed.

VIII. Acknowledgments

This research was supported by NASA under the NGATS-ATM Airspace Program (NNA06CN24A) and by NSF (ECCS-0745237). The authors are grateful to Marilyn Wolfson, Rich De Laura, Mike Matthews, Kim Calden, and Mike Robinson at MIT Lincoln Lab for help with CIWS data and fruitful discussions, and to John Robinson (NASA) for his interest in this work and insightful comments.

References

[1] Joint Economic Committee, Unites States Senate. Your flight has been delayed again. Technical report, May 2008.

[2] Bureau of Transportation Statistics. Understanding the reporting of causes of flight delays and cancellations. `http://www.bts.gov/help/aviation/html/understanding.html`, 2008.

[3] Joint Planning and Development Office. Next Generation Air Transportation System Integrated Plan, December 2004. `http://www.jpdo.aero/integrated_plan.html`.

[4] D. Bertsimas and S. Stock Patterson. The traffic flow management rerouting problem in air traffic control: A dynamic network flow approach. *Transportation Science*, 2000.

[5] P. K. Menon, G. D. Sweriduk, T. Lam, G. M. Diaz, and K. Bilimoria. Computer-aided Eulerian air traffic flow modeling and predictive control. *AIAA Journal of Guidance, Control and Dynamics*, 29:12–19, 2006.

[6] D. Sun, S. D. Yang, I. Strub, A. M. Bayen, B. Sridhar, and K. Sheth. Eulerian trilogy. In *Proceedings of the AIAA conference on Guidance, Navigation and Control*, 2006.

[7] J. Prete and J. S. B. Mitchell. Safe routing of multiple aircraft flows in the presence of time-varying weather data. In *AIAA Guidance, Navigation, and Control Conference and Exhibit, Providence, Rhode Island*, August 16-19 2004.

[8] J. Krozel, C. Lee, and J. S. B. Mitchell. Turn-constrained route planning for avoiding hazardous weather. *Air Traffic Control Quarterly*, 14(2):159–182, 2006.

[9] D. Bertsimas and A. Odoni. A critical survey of optimization models for tactical and strategic aspects of air traffic flow management. Technical report, NASA Ames Research Center, 1997.

[10] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization, 2007. Preprint.

[11] J. Krozel, J. S. B. Mitchell, V. Polishchuk, and J. Prete. Capacity estimation for airspaces with convective weather constraints. In *AIAA Guidance, Navigation and Control Conference and Exhibit, Hilton Head, South Carolina*, August 2007.

[12] J. S. B. Mitchell, V. Polishchuk, and J. Krozel. Airspace throughput analysis considering stochastic weather. In *AIAA Guidance, Navigation, and Control Conference, Keystone, Colorado*, 2006.

[13] R. DeLaura and S. Allan. Route selection decision support in convective weather: A case study. In *USA/Europe Air Traffic Management R&D Seminar, Budapest*, June 2003.

[14] D. Michalek and H. Balakrishnan. Building a stochastic terminal airspace capacity forecast from convective weather forecasts. In *Aviation, Range and Aerospace Meteorology Special Symposium on Weather-Air Traffic Management Integration, Phoenix, AZ*, January 2009.

[15] M. Wolfson, B. Forman, K. Calden, W. Dupree, R. Johnson, R. Boldi, C. Wilson, P. Bieringer, E. Mann, and J. Morgan. Tactical 0-2 hour convective weather forecasts for FAA. In *11th Conference on Aviation, Range and Aerospace Meteorology, Hyannis, MA*, 2004.

[16] Ravindra K. Ahuja, Thomas L. Magnanti, and James Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

[17] J. S. Mitchell. On maximum flows in polyhedral domains. In *SCG '88: Proceedings of the fourth annual symposium on Computational geometry*, pages 341–351, 1988.

[18] Georgia D. Tourassi, Erik D. Frederick, Mia K. Markey, and Jr. Carey E. Floyd. Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical Physics*, 28(12):2394–2402, 2001.

[19] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *Neural Networks, IEEE Transactions on*, 5(4):537–550, Jul 1994.

[20] Jason Van Hulse, Taghi M. Khoshgoftaar, and Amri Napolitano. Experimental perspectives on learning from imbalanced data. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 935–942, New York, NY, USA, 2007. ACM.

[21] Chao Chen, Andy Liaw, and Leo Breiman. Using random forest to learn imbalanced data. Technical report, Department of Statistics, UC Berkeley, 2004. `http://www.stat.berkeley.edu/users/chenchao/666.pdf`.

[22] Yang Liu, Aijun An, and Xiangji Huang. Boosting prediction accuracy on imbalanced datasets with SVM ensembles. In *Lecture Notes in Computer Science*, volume 3918. Springer Berlin / Heidelberg, 2006.

[23] Terry M. Therneau and Beth Atkinson. R port by Brian Ripley. *rpart: Recursive Partitioning*, 2008. R package version 3.1-42.

Author Biography

Diana Michalek is a Ph.D. candidate in the Operations Research Center at the Massachusetts Institute of Technology. She received her BAs in Mathematics and Computer Science at the University of California, Berkeley in May 2004. Afterwards she worked as a software engineer at Amazon.com in the Personalization and Digital Discovery teams. Her research interests are broadly in the applications of operations research, and include network flows, data mining, and air traffic flow management.

Hamsa Balakrishnan is an Assistant Professor of Aeronautics and Astronautics and of Engineering Systems at the Massachusetts Institute of Technology. She received her PhD in Aeronautics and Astronautics from Stanford University in April 2006, following which she was a researcher at the University of California, Santa Cruz and the NASA Ames Research Center. Her research interests include algorithms for the scheduling and routing of air traffic, techniques for the collection and processing of air traffic data, and mechanisms for the allocation of airport and airspace resources.